



Automatic Detection of Mild Cognitive Impairment from Spontaneous Speech using ASR

László Tóth¹, Gábor Gosztolya¹, Veronika Vincze¹, Ildikó Hoffmann^{2,3},
Gréta Szatlóczki⁴, Edit Biró⁴, Fruzsina Zsura⁴, Magdolna Pákáski⁴, János Kálmán⁴

¹ MTA-SZTE Research Group on Artificial Intelligence, Szeged, Hungary

² University of Szeged, Department of Linguistics, Szeged, Hungary

³ Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest, Hungary

⁴ University of Szeged, Department of Psychiatry, Szeged, Hungary

{ tothl, ggabor, vinczev } @inf.u-szeged.hu

Abstract

Mild Cognitive Impairment (MCI), sometimes regarded as a prodromal stage of Alzheimer’s disease, is a mental disorder that is difficult to diagnose. However, recent studies reported that MCI causes slight changes in the speech of the patient. Our starting point here is a study that found acoustic correlates of MCI, but extracted the proposed features manually. Here, we automate the extraction of the features by applying automatic speech recognition (ASR). Unlike earlier authors, we use ASR to extract only a phonetic level segmentation and annotation. While the phonetic output allows the calculation of features like the speech rate, it avoids the problems caused by the agrammatical speech frequently produced by the targeted patient group. Furthermore, as hesitation is the most important indicator of MCI, we take special care when handling filled pauses, which usually correspond to hesitation. Using the ASR-based features, we employ machine learning methods to separate the subjects with MCI from the control group. The classification results obtained with ASR-based feature extraction are just slightly worse than those got with the manual method. The F_1 value achieved (85.3) is very promising regarding the creation of an automated MCI screening application.

Index Terms: mild cognitive impairment, machine learning, temporal parameters of speech

1. Introduction

Alzheimer’s disease (AD) is a very distinct neurodegenerative disorder that develops for years before clinical manifestation. Although it has been extensively researched, uncertainty regarding its prodromal stages still exists. However, the symptoms of mild cognitive impairment (MCI) might be detected years before the actual diagnosis [1]. This tells us that the clinical appearance of AD is preceded by a prolonged, preclinical phase. Hence, precocious diagnosis and timely treatment are very important, as the progression can be decelerated and occurrence of new symptoms can be delayed [2].

MCI is a heterogenous syndrome that has clinical importance in the early detection of both AD [3] and the prodromal

state of dementia. MCI often remains undiagnosed, as recognizing cognitive impairment is challenging for clinicians at any stage of the disease: up to 50% of even later stage dementia fails to be recognized [4]. Widely-used screening tests such as the Mini-Mental State Examination (MMSE) are not sensitive enough to reliably detect subtle impairments present in patients with early MCI. Linguistic memory tests like word list and narrative recall are more effective in the detection of MCI, but they tend to produce undesired false positive diagnosis [5].

Although language impairment has been reported to occur precociously in the disease process [6], only cursory attention has been paid to a formal language evaluation when diagnosing AD [7]. Since language impairment has been reported even in the mild stage of AD, we recently developed a sensitive neuropsychological screening method that is based on a memory task, triggered by spontaneous speech [8]. In the future, this approach might permit the screening of MCI through a computerized, interactive test using a software package [9].

MCI is known to influence the (spontaneous) speech of the patient via three main aspects [10]. Firstly, the verbal fluency of the patient deteriorates, which results in distinctive acoustic changes – most importantly, in longer hesitations and a lower speech rate [5, 8, 11, 12]. Secondly, as the patient has trouble finding the right word, the lexical frequency of words and part-of-speech tags may also change significantly [13, 14, 15]. Thirdly, the emotional responsiveness of the patient was also observed to change in many cases. There are attempts to detect these changes based on the prosodic and paralinguistic features of the patient’s speech [16].

The MCI screening method we developed earlier focuses on the acoustic features [8]. We have shown experimentally that the proposed acoustic biomarkers indeed carry significant information for the separation of MCI patients from the control group. However, in this early study the transcription and annotation of speech signals was performed manually (with the help of the Praat software tool [17]). In this paper, we present our results in automatizing the biomarker extraction process using automatic speech recognition (ASR). In all the experiments, the manually extracted features of Hoffmann et al. [8] will serve as the baseline.

Other authors have also studied the acoustic correlates of MCI, and some also came up with automatic extraction methods. De Ipiña et al. applied the Praat tool to segment the utterance into voiced and voiceless sections [16]. Satt et al. also used Praat to discern voice/silent and periodic/apperiodic seg-

This publication is supported by the European Union and co-funded by the European Social Fund. Project title: Telemedicine-oriented research activities in the fields of mathematics, informatics and medical sciences. Project number: TÁMOP-4.2.2.A-11/1/KONV-2012-0073. Ildikó Hoffmann was supported by the Bolyai János Research Scholarship.

ments [11]. While these simple signal processing-based approaches can efficiently find silent pauses, the main problem with them is that they cannot detect filled pauses. Meanwhile, we found that about 10% of the hesitations in our database appear as filled pauses. Misclassifying these segments as speech can lead to an incorrect estimate of the amount of hesitation in the patient’s speech.

Lehr et al. used ASR to obtain the transcript of the signal, but they did not analyze acoustic features [18]. Fraser et al. also extracted acoustic features, but they used Nuance’s Dragon system instead of a dedicated ASR tool [19]. However, these out-of-box solutions do not help in finding and analyzing filled pauses. Roark et al. took care to annotate filled pauses, but they applied ASR only to force-align the manual annotation [5]. The study most similar to ours is that by Jarrold et al. [12]. They extracted both lexical transcripts and acoustic features using ASR. However, they used a standard word language model and did not appear to take special care with filled pauses.

Our targeted patient group tend to produce more grammatical errors and incorrectly inflected word forms. These errors would significantly increase the error rate of a standard ASR tool. However, we do not need precise word-level transcripts for the extraction of acoustic features like the speech rate and duration of pauses. Hence, we decided to train our ASR system to produce only phonetic-level transcripts. Moreover, we trained the system on a corpus of spontaneous speech where the filled pauses were explicitly annotated. The phone-level output of the recognizer allows us to extract features such as speech rate, while also allowing the collection of statistics about the duration of silent and filled pauses.

Based on the actual values of the acoustic indicators described above, in a second step a machine learning model is constructed, which seeks to decide whether a subject is likely to have MCI. We would like to add that we do not wish to diagnose the subjects, as this is the task of medically trained staff. Our goal here is to create an application that allows to perform a pre-filtering of the possible patients, which could then be followed by a diagnosis by a medical expert.

2. Indicators of MCI in Spontaneous Speech

Analyzing of the time course of speech has been shown to be an especially sensitive neuropsychological method for investigating cognitive processes such as speech production and planning [8]. Investigating the temporal parameters of spontaneous speech is vital because it can provide sensitive measures of a subject’s speech and language skills [20, 21].

In a study for Hungarian, the following parameters of speech were measured for AD patients and a normal control group: articulation rate, speech tempo, hesitation ratio, and rate of grammatical errors. The results showed that these parameters of speech may have a diagnostic value for mild-stage AD and therefore could be a useful aid in medical practice [8]. Other scientific studies have also confirmed that speech analysis could be a useful method in examining, or even diagnosing mild AD [5, 11, 12, 21, 1, 22]. In addition, lexical decision reaction time studies showed a longer overall latency in AD and MCI patients than in normal controls [23, 24, 25]. These results also confirm that speech analysis can contribute to the effective diagnosis of MCI.

In our earlier study on spontaneous speech in MCI the experimental setup for recording the utterances was as follows [8].

- | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ol style="list-style-type: none"> (1) Articulation rate was calculated as the number of phones per second during speech (excluding hesitations). (2) The speech tempo (phones per second) was calculated as the number of phones per second divided by the total duration of the utterance. (3) The length of utterance, given in milliseconds. (4-5) The duration of silent and filled pauses was calculated as the total duration of filled and silent pauses. (6-7) The number of silent and filled pauses reflects the absolute occurrence of silent and filled pauses, respectively. (8) The hesitation rate reflects the ratio of pauses and speech, which was calculated by dividing the length of the utterance by the total duration of pauses (both silent and filled). |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Table 1: A description of the eight acoustic biomarkers found to correlate with MCI by Hoffmann et al. [8].

After the presentation of a specially designed one-minute-long animated film, the subjects were asked to talk about the events seen on the film (immediate recall). After the presentation of a second film, the subjects were asked to talk about their previous day (spontaneous speech). As the last task, the subjects were asked to talk about the second film (delayed recall).

We measured the following acoustic parameters: articulation rate (1), speech tempo (2), length of utterance (3), duration of silent and filled pauses (hesitation) (4-5), number of silent and filled pauses (6-7) and hesitation rate (8). Hesitation was defined as the absence of speech for more than 30ms [26]. We should add that the absence of speech does not necessarily mean silence, but includes the filled pauses as well. Table 1 summarizes the eight acoustic indicators and how they were calculated.

3. Automatic Indicator Extraction using ASR

Calculating the above acoustic biomarkers manually (as was done in [8]) is quite expensive and requires skilled labor. Here we present our efforts towards the automatic extraction of the features of Table 1. One way of automation is to use signal processing methods. For example, Satt et al. employed the Praat software to segment the utterance into voice/silent and periodic/apertic parts [11]. However, these simple techniques cannot extract all the features of Table 1; for example, they cannot distinguish filled pauses from speech. The second option is to apply ASR. However, an off-the-shelf ASR tool (like the one used by Fraser et al. [19]) may be suboptimal. This is because standard speech recognizers are trained to minimize the transcription errors at the word level, while here we seek to extract non-verbal acoustic features like the rate of speech or the duration of silent and filled pauses. Note, for example, that none of the features in Table 1 require us to *identify* the phones; we need only to *count* them. Furthermore, while the filled pauses do not explicitly appear in the output of a standard ASR system, our feature set requires them to be found. And lastly, by examining the speech of dementia patients it was observed that the amount of agrammatical sentences and incorrect word in-

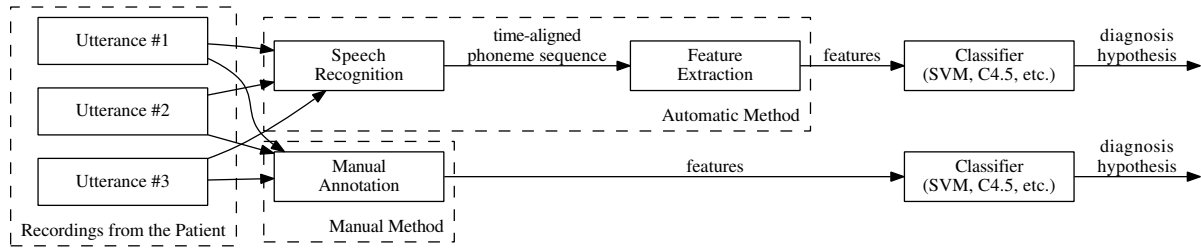


Figure 1: The steps of MCI detection using manual (lower path) or ASR-based (upper path) acoustic biomarker extraction.

flections increases [14]. It is practically impossible to prepare a standard ASR system to handle these errors. For these reasons we decided to use a speech recognizer that provides only a phone sequence as output (including filled pause as a special ‘phone’). Of course, recognizing the spontaneous speech of elderly people is known to be difficult [27]. Doing this without a vocabulary, only at the phonetic level clearly increases the number of errors. However, as we pointed out, not all types of phone recognition errors harm the extraction of our acoustic indicators. So the main question in the experiments was whether the acoustic indicators (and the subsequent classification step described in the next section) can tolerate the inaccuracies introduced by switching from manual to automatic extraction.

4. Classifying MCI

The overall goal of our project is to develop an application that would allow the user to self-test herself for MCI. Depending on the test results, the software would recommend that the subject visit a neurologist for a more thorough examination. We automated this decision making procedure using machine learning. In the experiments the values of the acoustic features were passed to the Weka toolkit [28], which classified the patient as either having MCI or not. The manually extracted feature values used by Hoffmann et al. in [8] were available for all the test files, and the classification results produced by Weka on this feature set served as our baseline. The feature extraction step was repeated using ASR, and then the resulting Weka scores were compared with the baseline. Fig. 1 compares the processing steps when using manual (lower path) or ASR-based (upper path) acoustic biomarker extraction.

5. Experimental Setup

5.1. ASR-based Biomarker Extraction

The speech recognizer was trained on the BEA Hungarian Spoken Language Database [29]. This database contains spontaneous speech, like the recordings collected from our MCI patients. We used roughly seven hours of speech data from the BEA corpus, mainly recordings from elderly persons, in order to match the age group of the targeted MCI audience. Although the BEA dataset contains spontaneous speech, its annotation did not quite suit our needs. It contained the word-level transcription of the utterances, but the filled pauses and other non-verbal audio segments (coughs, laughs, breath intakes, sighs etc.) were improperly marked. Hence we tailored the annotation of the recordings to our needs. This mainly consisted of adding filled pauses, breath intakes and exhales, laughter, coughs and gasps to the transcriptions in a consistent manner.

The ASR system was trained to recognize the phones in

the utterances, where the phone set included the special non-verbal labels listed above. For acoustic modeling we applied a special convolutional deep neural network-based technology. With this approach we managed to achieve one of the lowest phone recognition error rates on the TIMIT database [30]. As a language model we employed a simple phone bigram (again, including all the above-mentioned non-verbal audio tags).

The output of the ASR system is the phonetic segmentation and labeling of the input signal, which includes filled pauses. Based on this output, the acoustic biomarkers listed in Table 1 can be easily extracted using simple calculations.

5.2. MCI Classification

Our database of MCI patients is continuously growing; at the time of writing we had recordings taken from more than 100 persons. For various reasons (poor sound quality, controversial diagnosis, etc.) we had to filter out some patients, so in the experiments presented here we worked with the recordings of 51 subjects. From these 32 had MCI and 19 were control subjects, resulting in a 2-class classification task. For each subject we had three recordings for the three different tasks (for details on the tasks, see [8]). Using the eight biomarkers shown in Table 1, we got 24 features per patient. From a machine learning perspective, this is an extremely small dataset. However, the number of diagnosed MCI patients is limited, and collecting recordings of their speech is tedious. All the similar studies we found involved fewer than 100 patients [11, 12, 18, 5, 31].

Having so few examples, we did not create separate training and test sets, but applied the method of leave-one-out. That is, we withheld one example (i.e. one subject), trained our classifier on the remaining ones, and evaluated it on this withheld sample. We repeated this process for all the examples and then aggregated the results into one final score.

We used the Weka tool [28], which is a free, open-source collection of machine learning algorithms. Due to the small size of the dataset we restricted ourselves to simpler methods like linear SVM [32] and Random Forests [33]. Namely, we used the *SMO* and *RandomForest* algorithms of Weka. We optimized the parameter C of SVM as follows: we started from the default value (1.0), and doubled/halved it until the F -measure score for class MCI decreased twice in a row. We applied *RandomForest* with the default number of trees (100).

The choice of evaluation metric is not a clear-cut issue for this task. We can, of course, use standard Information Retrieval metrics: *precision* measures how many of the MCI hypotheses is for real occurrences, whereas *recall* tells us how many of the real MCI occurrences were detected. As there is evidently a trade-off between these two values, they are usually aggregated together by the F -measure (or F_1 -score), which is the harmonic

mean of precision and recall. However, as here we have a close-to-balanced class distribution, calculating the accuracy metric (defined as the number of correctly classified examples over the total number of examples) might make sense as well. Though we optimized the F_1 score of the MCI class, we list all four metrics in the tables.

5.3. Extending the Feature Set

The study that served as our starting point examined only the eight acoustic features shown in Table 1. The reason for this was that calculating and evaluating the features manually required an expensive workload. Here, however, we used an automatic method to get the time-aligned phoneme sequence of the utterances. Hence, we can readily extend the feature set by further features that can be calculated using the phone labels. Therefore, we looked for other features that we assumed could support the machine learning method applied in the second phase. This extended feature set was calculated as follows.

Firstly, we kept all the original features of Table 1. However, features (6) and (7) were altered slightly: instead of calculating the raw number of silent and filled pauses, we normalized them by dividing them by the total number of phones in the utterance. Furthermore, as we already have the length of each occurrence of silent/filled pauses, it was easy to extend the feature set with the mean and standard deviation of the lengths for these label occurrences. In addition, we observed that the ASR system often confused filled pauses with certain phones. For example, the most frequent sound uttered during hesitation is a schwa, which is easily confused with the vowel [ø]. Another example is substituting the hesitating word “hmm” with the phone [m]. Thus, we conjectured that an increase in the number and cumulative duration of these phones in the ASR output might indicate the presence of mis-recognized filled pauses. Hence we extended our feature set with features that describe the distribution of these phones in the utterance. More precisely, for the phones [m], [n] and [ø] we added the following four features to the feature set: cumulative duration (divided by the duration of the utterance), the number of occurrences (divided by the number of phonemes in the utterance), and the mean and standard deviation of the phone duration. With these extensions we obtained a set of 81 features, which will be referred to as the ‘extended’ feature set in the experiments.

Risk factors for MCI differ in men and women and are also known to vary with age [34]. These two attributes were also available for our training set, so we added them to the feature set, resulting in 26 and 83 features for the basic and extended feature sets, respectively. Of course, in the planned application we will not estimate these from the voice of the test subject, but the subject will be asked to provide this data when starting the test.

6. Results and Discussion

The results obtained can be seen in Table 2. Comparing the two classification methods, we see that SVM outperformed Random Forests with respect to all evaluation metrics except recall. As regards the feature sets, SVM performed best with the manually extracted feature set, achieving the highest values of F_1 and accuracy, while the precision and recall scores are also reasonably high (note that we optimized for F_1). By automatically extracting the features we got worse results, presumably due to the inaccuracies in the ASR output. However, with the extended feature set we achieved scores that are quite close to those with

Method	Feature set	Prec.	Recall	F_1	Acc.
SVM	Manual	82.4	87.5	86.2	82.4
	Automatic	83.9	81.3	82.5	78.4
	Extended	80.6	90.6	85.3	80.4
Random Forest	Manual	76.5	81.3	78.8	72.5
	Automatic	81.8	84.4	83.1	78.4
	Extended	76.3	90.6	82.9	76.5

Table 2: Results for the various classification methods and feature sets.

the manual feature set: the F_1 score of 85.3 is only slightly worse than the best manual value of 86.2. Notice also that the precision and recall scores are quite unbalanced in the case of the extended feature set. The gap could be decreased by adjusting the decision threshold, which would supposedly result in a higher F_1 score as well. Here, however, we tuned only the C parameter of SVM, mainly because only the later application will decide on the preferred balance of precision and recall.

With the Random Forest classifier, the results are somewhat mixed. For this classifier the extended feature set proved better or no worse than the manual one with respect to all evaluation metrics. Surprisingly, in this case the extended features performed no better than the simpler ‘automatic’ set (but the difference between the corresponding F_1 values of 82.9 and 83.1 is minimal). Although the recall value attained by Random Forest is the same as that for SVM (90.6), considering that all other scores are worse and that we optimized for F_1 , the scores overall clearly point in favor of using the SVM classifier.

Our results cannot be directly compared to those of others, as the database used was different. However, the diagnostic accuracies reported by other authors also fall in the 75%-90% range [5, 11]. Later practice will show if this score is sufficient for developing useful screening applications.

7. Conclusions

Mild cognitive impairment (MCI) is known to cause slight changes in the spontaneous speech of the patient. Our starting point was a study that found eight acoustic correlates of MCI, but applied a manual method for the extraction of these features from the sound files. In this study, we sought to automate the feature extraction process by applying ASR. Unlike earlier authors, we used ASR to extract only a phonetic level segmentation and annotation. Furthermore, we took special care with filled pauses, which correspond to hesitations in most cases. We also extended the originally proposed features with further ones we considered informative. In the second step, using these features, we employed simple machine learning methods to separate the subjects with MCI from the control subjects. Our results showed that by switching from the manual to the ASR-based feature extraction method the F_1 score decreased only slightly. The F_1 value we got (85.3) is very promising regarding the creation of an automated MCI screening application.

While in this study we analyzed only acoustic features, it is known that the linguistic content of the speech can also be used to detect MCI or the early stage of Alzheimer’s disease [7]. Some authors have already made steps towards automating the linguistic analysis part using ASR [14, 15, 18]. We also have some preliminary results in this direction, and we plan to combine the acoustic and linguistic analysis methods in the future.

8. References

- [1] K. L. de Ipiña, J.-B. Alonso, C. M. Travieso, J. Sol-Casals, H. Eguiraun, M. Faundez-Zanuy, A. Ezeiza, N. Barroso, M. Ecay-Torres, P. Martínez-Lage, and U. M. de Lizardui, "On the selection of non-invasive methods based on speech analysis oriented to automatic Alzheimer disease diagnosis," *Sensors*, vol. 13, no. 5, pp. 6730–6745, 2013.
- [2] J. Kálmán, M. Pákáski, I. Hoffmann, G. Drótos, G. Darvas, K. Boda, T. Bencsik, A. Gyimesi, Z. Gulyás, M. Bálint *et al.*, "Early mental test – developing a screening test for mild cognitive impairment," *Ideggyógyászati szemle*, vol. 66, no. 1-2, pp. 43–52, 2013.
- [3] S. Negash, L. Petersen, Y. Geda, D. Knopman, B. Boeve, G. Smith, R. Ivnik, D. Howard, J. Howard Jr, and R. Petersen, "Effects of ApoE genotype and Mild Cognitive Impairment on implicit learning," *Neurobiology of Aging*, vol. 28, no. 6, pp. 885–893, 2007.
- [4] L. Boise, M. Neal, and J. Kaye, "Dementia assessment in primary care: Results from a study in three managed care systems," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 59, no. 6, pp. M621–M626, 2004.
- [5] B. Roark, M. Mitchell, J.-P. Hosom, K. Hollingshead, and J. Kaye, "Spoken language derived measures for detecting mild cognitive impairment," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2081–2090, 2011.
- [6] A. P. Association, *DSM-IV-TR*. American Psychiatric Association, 2000.
- [7] K. Bayles, "Language function in senile dementia," *Brain and Language*, vol. 16, no. 2, pp. 265–280, 1982.
- [8] I. Hoffmann, D. Németh, C. Dye, M. Pákáski, T. Irinyi, and J. Kálmán, "Temporal parameters of spontaneous speech in alzheimer's disease," *International Journal of Speech-Language Pathology*, vol. 12, no. 1, pp. 29–34, 2010.
- [9] J. Kálmán, I. Hoffmann, A. Hegyi, G. Drótos, A. Heilmann, and M. Pákáski, "Spontaneous speech based web screening test for MCI," in *Proceedings of ADI*, San Juan, Puerto Rico, 2014, pp. 315–318.
- [10] C. Laske, H. R. Sohrabi, S. M. Frost, K. L. de Ipiña, P. Garrard, M. Buscema, J. Dauwels, S. R. Soekadar, S. Mueller, C. Linneemann, S. A. Bridenbaugh, Y. Kanagasingam, R. N. Martins, and S. E. O'Bryant, "Innovative diagnostic tools for early detection of Alzheimer's disease (in press)," *Alzheimer's & Dementia*, 2015.
- [11] A. Satt, R. Hoory, A. König, P. Aalten, and P. H. Robert, "Speech-based automatic and robust detection of very early dementia," in *Proceedings of Interspeech*, Singapore, 2014, pp. 2538–2542.
- [12] W. Jarrold, B. Peintner, D. Wilkins, D. Vergryi, C. Richey, M. L. Gorno-Tempini, and J. Ogar, "Aided diagnosis of dementia type through computer-based analysis of spontaneous speech," in *Proceedings of CLPsych*, Baltimore, Maryland, USA, 2014, pp. 27–37.
- [13] V. Baldas, C. Lampiris, C. Capsalis, and D. Koutsouris, "Early diagnosis of Alzheimers type dementia using continuous speech recognition," in *Proceedings of MobiHealth*, Ayia Napa, Cyprus, 2011, pp. 105–110.
- [14] K. C. Fraser, J. A. Meltzer, N. L. Graham, C. Leonard, G. Hirst, S. E. Black, and E. Rochon, "Automated classification of primary progressive aphasia subtypes from narrative speech transcripts," *Cortex*, vol. 55, pp. 43–60, 2014.
- [15] P. Garrard, V. Rentoumi, B. Gesierich, B. Miller, and M. L. Gorno-Tempini, "Machine learning approaches to diagnosis and laterality effects in semantic dementia discourse," *Cortex*, vol. 55, pp. 122–129, 2014.
- [16] K. L. de Ipiña, J. B. Alonso, J. Solé-Casals, N. Barroso, P. Henriquez, M. Faundez-Zanuy, C. M. Travieso, M. Ecay-Torres, P. Martínez-Lage, and H. Eguiraun, "On automatic diagnosis of Alzheimer's disease based on spontaneous speech analysis and emotional temperature," *Cognitive Computation*, vol. 7, no. 1, pp. 44–55, 2015.
- [17] P. Boersma, "Praat, a system for doing phonetics by computer," *Glott international*, vol. 5, no. 9/10, pp. 341–345, 2002.
- [18] M. Lehr, E. Prudhommeaux, I. Shafran, and B. Roark, "Fully automated neuropsychological assessment for detecting Mild Cognitive Impairment," in *Proceedings of Interspeech*, Portland, OR, USA, 2012.
- [19] K. Fraser, F. Rudzicz, N. Graham, and E. Rochon, "Automatic speech recognition in the diagnosis of primary progressive aphasia," in *Proceedings of SLPAT*, Grenoble, France, 2013, pp. 47–54.
- [20] S. Baum, S. Blumstein, M. Naeser, and C. Palumbo, "Temporal dimensions of consonant and vowel production: An acoustic and CT scan analysis of aphasic speech," *Brain and Language*, vol. 39, no. 1, pp. 33–56, 1990.
- [21] J. Illes, "Neurolinguistic features of spontaneous language production dissociate three forms of neurodegenerative disease: Alzheimer's, Huntington's, and Parkinson's," *Brain and language*, vol. 37, no. 4, pp. 628–642, 1989.
- [22] J. Meilán, F. Martínez-Sánchez, J. Carro, D. López, L. Millian-Morell, and J. Arana, "Speech in Alzheimer's disease: can temporal and acoustic parameters discriminate dementia?" *Dementia and Geriatric Cognitive Disorders*, vol. 37, no. 5-6, pp. 327–334, 2014.
- [23] V. Taler and G. Jarema, "On-line lexical processing in AD and MCI: An early measure of cognitive impairment?" *Journal of neurolinguistics*, vol. 19, no. 1, pp. 38–55, 2006.
- [24] F. Cuetos, T. Martínez, C. Martínez, C. Izura, and A. Ellis, "Lexical processing in Spanish patients with probable Alzheimers disease," *Cognitive Brain Research*, vol. 17, no. 3, pp. 549–561, 2003.
- [25] P. Walla, E. Püregger, J. Lehrner, D. Mayer, L. Deecke, and P. Dal Bianco, "Depth of word processing in Alzheimer patients and normal controls: a magnetoencephalographic (MEG) study," *Journal of Neural Transmission*, vol. 112, no. 5, pp. 713–730, 2005.
- [26] M. Gósy, "The paradox of speech planning and production (in Hungarian)," *Magyar Nyelvőr*, vol. 12, no. 1, pp. 3–15, 1998.
- [27] B. Ramabhadran, J. Huang, and M. Picheny, "Towards automatic transcription of large spoken archives – english ASR for the MALACH project," in *Proceedings of ICASSP*, 2003, pp. 216–219.
- [28] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, "The WEKA data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
- [29] M. Gósy, "BEA: A multifunctional Hungarian spoken language database," *The Phonetician*, vol. 105, no. 106, pp. 50–61, 2012.
- [30] L. Tóth, "Convolutional deep maxout networks for phone recognition," in *Proceedings of Interspeech*, 2014, pp. 1078–1082.
- [31] K. C. Fraser, F. Rudzicz, and E. Rochon, "Using text and acoustic features to diagnose progressive aphasia and its subtypes," in *Proceedings of Interspeech*, Lyon, France, 2013, pp. 25–29.
- [32] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.
- [33] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [34] P. Sachdev, D. Lipnicki, J. Crawford, S. Reppermund, N. Kochan, J. Trollor, B. Draper, M. Slavin, K. Kang, O. Lux, and K. M. H. Brodaty, "Risk profiles for mild cognitive impairment vary by age and sex: the sydney memory and ageing study," *The American Journal of Geriatric Psychiatry*, vol. 20, no. 10, pp. 854–865, 2012.